DOCUMENT RESUME

ED 230 571 TM 830 252

AUTHOR Daniel, Mark
TITLE A Factorial Study of Reasoning Tests. Technical Report 1982-6.
INSTITUTION Johnson O'Connor Research Foundation, Chicago, IL. Human Engineering Lab.
PUB DATE Sep 82
NOTE 27p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Abstract Reasoning; *Cognitive Tests; *Correlation; *Factor Analysis; Group Testing; Higher Education; *Induction; *Logical Thinking; Pattern Recognition; Reaction Time; Secondary Education; Test Validity
IDENTIFIERS *Reasoning Tests

ABSTRACT

The Inductive Reasoning and Analytical Reasoning tests of the Johnson O'Connor Research Foundation have been analyzed in a factorial study including marker tests for two commonly recognized reasoning factors, induction (three tests) and logical reasoning (two tests). The primary purpose of the study was to determine the extent to which the Johnson O'Connor tests measure these factors. Multiple groups factor analysis showed the two factors to be highly correlated (r=.73) and practically indistinguishable; a single "general" reasoning factor defined by all five marker tests does a good job of explaining their intercorrelations. The Analytical Reasoning test loads highly on this general factor, but Inductive Reasoning (individually administered) is highly independent of the general factor. It appears that the Inductive Reasoning test measures a distinct reasoning aptitude characterized by quickness in seeing common elements. The group-administered Inductive Reasoning test has low reliability and does not correlate highly with the individual version. (Author)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED230571

# A FACTORIAL STUDY OF REASONING TESTS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Baker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Mark Daniel

JOHNSON O'CONNOR RESEARCH FOUNDATION

HUMAN ENGINEERING LABORATORY

Technical Report 1982-6

September, 1982

TM 830 252

ACKNOWLEDGMENTS

Our thanks are due to the Educational Testing Service, Institute for Personality and Ability Testing, and Sheridan Psychological Services for permission to use their tests in this research project. We would also like to thank Robert Demaree of Texas Christian University for kindly reviewing the proposal for this project and for suggesting certain methods of data analysis.

CONTENTS


| | Page |
|---|---|
| Acknowledgments | i |
| Table of Contents | ii |
| List of Tables | iii |
| Abstract | 1 |
| Earlier Factorial Studies of Reasoning | 1 |
| Description of Johnson O'Connor Research Foundation Reasoning Tests | 3 |
| Research Questions Addressed in This Study | 3 |
| Procedures | 5 |
| I. Variables | 5 |
| II. Subjects | 6 |
| III. Methods of Data Collection | 7 |
| IV. Methods of Analysis | 7 |
| V. Treatment of Extraneous Variables | 8 |
| Results | 9 |
| I. Descriptive Statistics | 9 |
| II. Reliability of Group Inductive Reasoning | 11 |
| III. Examination of Hypothesized Factors | 11 |
| IV. Overlap of JOC Reasoning Tests with Reasoning Factors | 14 |
| V. Difference Between the Two Inductive Reasoning Tests | 14 |
| VI. Relationship of Reasoning Tests to Other Aptitudes | 17 |
| Conclusions | 19 |
| References | 20 |

LIST OF TABLES


| | | Page |
|---|---|---|
| 1. | Descriptive Statistics for Reasoning Tests (N=615) | 10 |
| 2. | Intercorrelations of Reasoning Tests (N=615) | 10 |
| 3. | Factor Structure Matrix: Multiple-Groups Factor Analysis | 11 |
| 4. | Mean Correlations Between Marker Tests Defining the Same Factor or Different Factors | 11 |
| 5. | Loadings on a General Factor Defined by the Five Marker Tests | 12 |
| 6. | Residuals After Extraction of One (Above Diagonal) or Two (Below Diagonal) Multiple-Group Factors | 12 |
| 7. | Loadings on Diagonal Factor Defined by Inductive Reasoning, Wks. 164 | 13 |
| 8. | Results of Canonical Factor Analysis | 14 |
| 9. | Multiple Regressions of Multiple-Group Factors on JOC Reasoning Tests | 14 |
| 10. | Multiple Regressions of Inductive Reasoning Difference Scores (700-164) on Other Aptitude Measures | 16 |
| 11. | Correlations of Reasoning Tests with Other Aptitude Tests (Males/Females) | 18 |

# A Factorial Study of Reasoning Tests

Mark Daniel

*Abstract: The Inductive Reasoning and Analytical Reasoning tests of the Johnson O'Connor Research Foundation have been analyzed in a factorial study including marker tests for two commonly recognized reasoning factors, induction (three tests) and logical reasoning (two tests). The primary purpose of the study was to determine the extent to which the Johnson O'Connor tests measure these factors. Multiple groups factor analysis showed the two factors to be highly correlated (r = .73) and practically indistinguishable; a single "general" reasoning factor defined by all five marker tests does a good job of explaining their intercorrelations. The Analytical Reasoning test loads highly on this general factor, but Inductive Reasoning (individually administered) is highly independent of the general factor. It appears that the Inductive Reasoning test measures a distinct reasoning aptitude characterized by quickness in seeing common elements. The group-administered Inductive Reasoning test has low reliability and does not correlate highly with the individual version.*

The Inductive Reasoning test (Worksample 164) and the Analytical Reasoning test (Worksample 244) have been administered by the Johnson O'Connor Research Foundation for over forty years. Briefly, the former test measures speed in picking three pictures that have something in common out of a row of six pictures, and the latter measures the ability to arrange a set of words so that they form a logical pattern. Occupational and educational validation studies show that lawyers, physicians, writers and editors tend to score high on the Inductive Reasoning test, and editors, computer programmers and actuaries obtain high scores on Analytical Reasoning. The Analytical Reasoning test has a substantial relationship with high school grades (especially in mathematics) and with SAT-Math scores, whereas Inductive Reasoning is almost completely unrelated to academic criteria. A third test, group-administered Inductive Reasoning (Worksample 700), has recently been developed in the hope that it might serve as an alternate form to Worksample 164. As yet, nothing is known about the reliability or validity of this new test.

The present research is a construct-validation study of these three reasoning tests. The primary purpose is to determine the extent to which they measure the same reasoning factors that have been identified by other investigators, using a wide variety of tests, over the past fifty years. A secondary purpose is to evaluate the group-administered Inductive Reasoning test, both for its usefulness as an alternate form and for its potential as a measure of other types of reasoning.

## EARLIER FACTORIAL STUDIES OF REASONING

The following historical review is based on a report by Northrup (1977). In a series of factor analytic studies, Thurstone (1938; Thurstone & Thurstone, 1941) identified one to three somewhat distinct reasoning factors.

The clearest of these was labeled "induction" (I) because the tests which defined the factor all required that a general rule be inferred from a set of examples. Some of the tests having high loadings on this factor are: Figure Classification, in which two or more sets of figures are presented, and the examinee must discern the rule by which figures can be assigned to the sets; Number Series; Marks (Locations), in which the examinee must infer from several examples the rule governing the placement of a mark in a row of dashes and spaces; Tabular Completion, in which missing entries in a table of data must be filled in; Letter Series; and Letter Groupings (Letter Sets). Two other factors were less well defined and were not always found in factor analyses. One, referred to as "restrictive reasoning" because the defining tests involve some restriction in the solution, had high loadings for Arithmetic Reasoning and Numerical Judgment tests. The other, labeled "deduction," is defined by tests of syllogistic reasoning such as False Premises (Nonsense Syllogisms).

As mentioned, the studies were not consistent in the reasoning dimensions they produced. For example, the Arithmetic Reasoning test loaded at various times on the induction, restrictive reasoning and deduction factors. Such variability in results is partly due to differences in the composition of the test batteries used. However, it also reflects the fact that any reasoning task requires the application of more than one factor of reasoning. For example, on an induction test one uses deduction when applying an inferred rule to the examples to see if it fits. For this reason, it is to be expected that reasoning factors will be moderately correlated with one another.

Later factor analytic investigations of reasoning tests were conducted by a number of researchers. During the Second World War the Army Air Force studied many tests and tended to find a "general reasoning" factor defined by such tests as Arithmetic Reasoning, Spatial Reasoning (a variant of the Marks or Locations test), Figure Analogies and Number Series (Guilford & Lacey, 1947). However, these experiments included few or none of the tests that had previously been found to define the induction or deduction factors. Adkins and Lyerly (1952) found a deduction factor with high loadings for False Premises and Verbal Analogies, and three induction factors which differ in their emphasis on hypothesis formulation vs. hypothesis testing. They grouped three Figure Classification tests with a factor of "flexibility of perceptual closure," based on research by Pemberton (1952) showing a relationship between these two types of tests. Research by Guilford and his associates (Michael, Zimmerman & Guilford, 1951; Green, Guilford, Christensen & Comrey, 1953; Guilford, Green, Christensen, Hertzka & Kettner, 1954) tended to support the existence of three reasoning factors: general reasoning (arithmetic reasoning), logical reasoning (deduction), and induction.

Analyses conducted over the past twenty years have in general confirmed the categorization of reasoning factors presented by Guilford. The ETS Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harman, 1976) provides marker tests for the three factors, which it describes as follows:

General Reasoning: "The ability to select and organize relevant information for the solution of a problem." (p. 133)

Logical Reasoning: "The ability to reason from premise to conclusion, or to evaluate the correctness of a conclusion." (p. 141)

Induction: "This factor identifies the kinds of reasoning abilities involved in forming and trying out hypotheses that will fit a set of data." (p. 79)

Ekstrom et al. point out that induction may include several sub-abilities, one of which would be that measured by the Figure Classification test. They also note that logical and general reasoning cannot always be clearly distinguished; both involve the serial processing of information, but the type of information differs (numbers vs. words).

## DESCRIPTION OF JOHNSON O'CONNOR RESEARCH FOUNDATION REASONING TESTS

Inductive Reasoning, Worksample 164: This is a measure of the speed of seeing a common element in three out of six pictures. The test is individually administered. Each of the 58 items has its own four-point scoring scale based on completion time; incorrect responses (which are infrequent) are given the lowest score value. The test is highly speeded: the average item has a median completion time of about seven seconds. Despite this fact, the test's correlation with perceptual speed (measured by Number Checking) is very low, and the correlation with English vocabulary knowledge is also below .2 (Daniel, 1978). The internal-consistency reliability of the test is .88 (Daniel, 1979).

Analytical Reasoning, Worksample 244: In this test, the examinee must arrange a set of words in a pre-determined pattern so that the words form a logical or temporal sequence. The test is individually administered, and as on Wks. 164 each item is scored according to completion time; however, the test is not highly speeded (median completion times on the thirteen items range from five to fifty-three seconds). Internal-consistency reliability is .65 (Daniel & Kling, 1974). The test has moderate (about .4) correlations with English vocabulary, spatial visualization, and the Inductive Reasoning test. A recent factor analysis (Trafton, 1981) found Analytical Reasoning to load .56 on a logical reasoning factor defined by marker tests from the ETS Kit, .46 on a spatial visualization factor, and .38 on a factor defined by Worksample 164.

Inductive Reasoning, Worksample 700: The thirty items on this test are similar to those of Worksample 164, but the test is group administered, so that individual items cannot be timed. Instead, the items are administered in groups of ten with a very short time limit for each group. This test is new and has not previously been evaluated for reliability or correlations.

## RESEARCH QUESTIONS ADDRESSED IN THIS STUDY

1) How are each of the three Johnson O'Connor (JOC) reasoning tests related to the induction and logical reasoning factors? A major purpose of

this research project is to tie these tests to established constructs familiar to other researchers. The general reasoning factor will not be studied because none of the JOC reasoning tests have numerical content.

It is expected that the Analytical Reasoning test will, as before, have a moderately high loading on the logical reasoning factor. Although ordering words into a logical or temporal sequence is not the same mental process as evaluating statements to see whether they can be logically deduced from a set of premises, both tasks seem to involve logical, step-by-step thinking. It is also hypothesized that the Analytical Reasoning test will have a low loading on the induction factor, because the test does not appear to require the examinee to infer rules.

The Inductive Reasoning tests may have less in common with the induction factor than the similarity of their names would suggest. In particular, the highly speeded nature of Worksample 164 may cause it to diverge from both of the established reasoning constructs. This test seems to be measuring quickness in seeing relationships; the hypothesis-testing aspect of the test (and of Worksample 700) appears to be much less important than on the marker tests for induction. A low to moderate loading is expected for the Inductive Reasoning tests on the induction factor. Because there is no deductive, step-by-step reasoning involved in the Inductive Reasoning tests, their loadings on the logical reasoning factor should be low.

If almost all of the reliable variance of any of the JOC reasoning tests is shown to overlap with the induction and logical reasoning factors, then we will be in a position to interpret those tests in terms of known constructs. In particular, if any test correlates highly with one factor, it can be considered to be a new marker test for that factor, and validation evidence which applies to the factor can also be applied to the test (and vice-versa). If, on the other hand, a substantial proportion of the reliable variance of a JOC reasoning test is not explained by the two reasoning factors, nor by other ability factors included in the analysis, then we will have evidence for a new dimension of reasoning ability.

2) What is the correlation between the induction and logical reasoning factors? As mentioned previously, tests in the reasoning domain often cannot be clearly assigned to one or the other of the established reasoning factors. Instead, many tests appear to include several types of reasoning. This suggests that the reasoning factors are themselves correlated.

3) How are the two reasoning factors related to the JOC reasoning tests? For the reasons mentioned above, the induction factor may be only moderately related to the JOC tests. If this is the case, then the JOC battery fails to measure a well-known ability dimension, and consideration should be given to adding a test of induction to the battery.

4) How similar are the two Inductive Reasoning tests? These tests have similar item type but different forms of administration. On Worksample 164 the total score primarily reflects the examinee's average speed; errors are infrequent and usually have little effect on the score. On Worksample 700, the total score is the sum of the number of items answered correctly in each group of ten. Although it is hoped that the short time limits for

the item groups will cause the test to measure primarily speed, power may have more of an influence on this test than on Worksample 164, because an examinee who gets stuck on one of the early or middle items of a group could obtain a low score for that group despite a high rate of speed for most items attempted. Of interest will be the intercorrelation of Worksamples 164 and 700 and a comparison of their correlations with the reasoning and other ability factors present in the battery.

If Worksample 164 is strongly related to the induction and logical reasoning factors, then the usefulness of Worksample 700 can be judged directly from its own loadings on the two reasoning factors. If, however, Worksample 164 does not have substantial overlap with the reasoning factors, then the comparability of the two Inductive Reasoning tests will probably be the question of primary interest.

## PROCEDURES

I. Variables. The three JOC reasoning tests have already been described. The marker tests for the induction factor are:

Figure Classification (ETS Kit of Factor-Referenced Cognitive Tests, Ekstrom et al., 1976). Each item presents two or three groups of geometric figures; the examinee must figure out the rule by which each of 8 test figures can be assigned to one of the groups. There are fourteen items, and the time limit is 8 minutes. Alternate-form reliability is about .64.

Locations (ETS Kit). Each of the 13 items shows five lines of dashes and spaces. In each of the first four lines, one of the dashes is replaced by an X. The examinee must figure out the location rule and decide where the X should be placed in the fifth line. The time limit is 6 minutes. Alternate-form reliability is about .60. (Note: Item 5 was not administered because it was judged to be much too difficult for its location in the test.)

Letter Sets (Comprehensive Ability Battery, Hakstian & Cattell, 1976). Each of the 12 items has five groups of letters, with four letters per group. The examinee must identify one group of letters which does not follow the rule or pattern shown by the other four groups. The time limit is 6 minutes. Split-half reliability is about .74.

The marker tests for the logical reasoning factor are:

Inference (ETS Kit). There are ten items. Each presents a set of statements, and the examinee chooses which one of five possible conclusions may be drawn from the statements. The time limit is 6 minutes. Alternate-form reliability is about .59.

Logical Reasoning (Hertzka & Guilford, 1955). This test has 20 items. It differs from the Inference test in that the items, while verbal, require more formal syllogistic reasoning. The time limit is 10 minutes. Split-half reliability is about .80.

These marker tests were selected because of their demonstrated relationship to the two reasoning factors of interest. The Letter Sets and Logical Reasoning tests are similar to tests contained in the ETS Kit. All of the marker tests except Letter Sets have two parts; only the first part of each test was administered in this study, because of the limited amount of testing time available. The reliabilities and other information given above are for the first part only.

The examinees participating in this study also took the entire Johnson O'Connor Research Foundation battery of aptitude and knowledge tests. The following tests from that battery will be included in some of the analyses because they measure abilities which may be related to the reasoning tests. These tests are:

Number Checking (Wks. 613): speed and accuracy in comparing pairs of numbers.
Ideaphoria (Wks. 161): quantity written in response to an open-ended question.
Wiggly Block (Wks. 4): a three-dimensional assembly test of structural visualization.
Paper Folding (Wks. 622): another spatial test, similar to Thurstone's Punched Holes test.
Memory for Design (Wks. 294): accuracy in reproducing an abstract straight-line drawing from memory.
Silograms (Wks. 376): paired-associate memory, using pairs of nonsense words and English words.
Number Memory (Wks. 165): ability to remember eight six-digit numbers.
Object Observation (Wks. 206): noticing small changes in a picture.
Writing Speed (Wks. 301).
English Vocabulary (Wks. 690): a recognition test of words in general usage.

Thus, the Johnson O'Connor test battery measures the broad aptitudes of perceptual speed, ideational fluency, spatial visualization, memory, and vocabulary.

II. Subjects. The examinees in this study are clients of the Johnson O'Connor Research Foundation, who were tested in all fourteen of the Foundation's offices around the country during the fall of 1981. These people came to the Foundation for testing in order to obtain information about their aptitudes which they could use in career and educational planning. They either paid a fee for the testing, or had their fee paid under an independent grant supporting the testing of high school and college students. While the age of the examinees taking the reasoning marker tests ranges from 13 to 57, 91% of them are between the ages of 16 and 40, and this more restricted age range has been used in order to reduce age effects on test scores. A total of 615 examinees aged 16-40 (363 males and 252 females) have complete data on all eight of the reasoning tests.

The examinee sample is relatively homogeneous with respect to several demographic characteristics. The level of education is fairly high: those aged 25 and older have completed an average of 15.9 years of schooling, and the great majority of those in high school plan to attend college.

Fathers of examinees have completed an average of 16.0 years of education (Bowker & Moore, 1981). No statistics on family income are available, but most examinees are in the middle or upper-middle income groups. Very few examinees are from racial minority groups.

III. Methods of Data Collection. The marker tests for reasoning were administered as if they were part of the standard JOC test battery. The Locations test was administered during a session of individually-administered tests; the other four marker tests were given to small groups (fewer than ten examinees at a time) using tape-recorded instructions and timing. On all of the marker tests, the instruction pages were revised to contain only the practice items; the instructions themselves were given verbally. Also, on the tests from the ETS Kit, the instructions were changed so as to tell the examinee that his score would be the number of items answered correctly, rather than the number right minus a fraction of the number wrong. These changes were made so that the marker tests would resemble the JOC tests, which do not require the examinee to read written instructions and which encourage guessing. Aside from these changes in format, the administration and timing procedures specified by the test publishers were followed. After all of the marker tests had been administered, examinees were informed that these tests were experimental, and the nature of the research project was briefly described.

IV. Methods of Analysis. To address the first and second research questions (i.e., how the JOC reasoning tests are related to the established reasoning factors, and what the correlation is between the reasoning factors) a multiple-group common factor analysis will be performed on the eight reasoning tests (Gorsuch, 1974). This is a confirmatory type of analysis, in which the expected factors are specified ahead of time in terms of the variables (tests) that will define them. In this instance, two of the marker tests (Inference and Logical Reasoning) will define the logical reasoning factor, and the other three marker tests (Locations, Figure Classification and Letter Sets) will define the induction factor. The JOC reasoning tests will have no influence on the structure of these factors; however, their correlations with the factors can be determined. A common factor analysis (which focuses on shared variance among the tests) is more appropriate for this study than a component analysis (which considers both common variance and variance unique to individual tests), because the theoretical factors of interest are determined by what the marker tests have in common. Test reliabilities will be used in the diagonal of the correlation matrix as communality estimates.

The correlation between the induction and logical reasoning factors, and the correlations of the JOC reasoning tests with the factors, will be calculated. The correlations of the marker tests with the two factors will also be examined to check the accuracy of the assumption that these tests measure two distinct (although correlated) factors. The residual correlation matrix, after extraction of the first two reasoning factors, will be examined and if it appears that a substantial amount of correlation remains, one or more additional factors will be derived.

Following the multiple-groups analysis, a second check on the factor structure of the reasoning tests will be made using a maximum-likelihood

common factor technique (Rao's canonical factor analysis). This analysis is particularly useful because it offers an objective statistical method for determining the number of factors needed to explain the correlation matrix. This analysis will be performed on both the marker and JOC tests, and a two-factor solution will be imposed; we can then evaluate how well the empirically-derived factors agree with the anticipated factors, and can see whether two factors are sufficient to explain the correlations among all eight reasoning tests.

The third research question (i.e., how well the JOC battery measures the induction and logical reasoning factors) will be addressed through multiple correlation of each of the factors with the JOC reasoning tests, based on the results of the multiple groups analysis described above. Of primary interest will be the extent to which the JOC reasoning tests can predict the reasoning factors.

The fourth research question concerns the similarity of the two Inductive Reasoning worksamples. A direct measure of the extent to which they measure the same ability is their intercorrelation corrected for test unreliability. Unless this correlation is very high (above .9), it will also be desirable to describe the difference between the two tests in terms of other aptitudes. For this purpose, the difference between standardized scores on Wks. 700 and Wks. 164 will be used as the dependent variable in a multiple regression analysis, with other ability tests as predictors. Any differences between these tests will also be illuminated by comparison of their loadings on the induction and logical reasoning factors from the initial multiple-groups factor analysis.

Finally, it will be of interest to evaluate all of the reasoning tests in terms of their correlations with other abilities measured in the JOC test battery, such as perceptual speed, structural visualization, memory, and English vocabulary. This analysis will show whether some tests are more independent of external factors than others; there is reason to expect the marker tests of logical reasoning, for example, to correlate substantially with English vocabulary, and the Figure Classification test may be significantly influenced by spatial visualization or design memory. Examination of correlations with other aptitudes may shed light on the intercorrelations among the reasoning tests, and also may be useful in guiding the selection of tests to be used in a battery.

V. Treatment of Extraneous Variables. Age is related to scores on many of the tests in this study, and there are sex differences on several tests (e.g. Wiggly Block, Number Checking, Silograms). Most of the JOC tests are scored with centile age norms, and the use of these centile scores rather than raw scores will remove age-related score variance. On tests for which age norms are not available, a check will be made for linear and quadratic relationships between age and raw score. If on any of these tests age variation accounts for more than one percent of test variance, it will be removed by subtracting the predicted test score (based on the linear or quadratic regression equation) from the observed test score.

The proper method for treating age-related test score variance in factor analysis is controversial (Horn & Cattell, 1982), and cogent arguments can

be made in favor of either controlling or not controlling for age effects in the present study. To leave age uncontrolled would be to allow the similarity or difference between the age curves of two tests to affect their correlation. This may be desirable from the point of view that a difference between two age curves is *prima facie* evidence that the two tests are measuring different abilities. On the other hand, variation in test scores with age sometimes may reflect the influence of extraneous factors (such as speed) on test performance; two tests which measure the same ability may have their intercorrelation reduced if one is much more highly speeded than the other. Also, if age is not controlled, the results of a correlational study will vary with the age range of the sample of subjects. Controlling for age, as in the present study, gives results which should be similar to those found in a study using subjects in a narrow age range.

Sex differences present a similar problem for analysis. Among the abilities measured in the present study, females are known to score higher than males in graphoria (number-checking speed), memory for nonsense words, memory for visual detail, and writing speed, and males score higher in spatial visualization. As with age, the fact that a sex difference appears on one test but does not appear (or appears in the opposite direction) on another is, in itself, evidence of a factorial difference between the tests. If sex differences are not controlled, the similarity or difference between sex differences on two tests will respectively raise or lower their intercorrelation. It is common practice in research to remove sex-related test score variance in some way, either by standardizing scores separately within sex, by holding sex statistically constant, or by analyzing the data separately for males and females. Because there are substantial numbers of male and female cases in the present data set, analyses will be conducted separately for the two sexes whenever tests showing sex differences are involved.

## RESULTS

I. Descriptive Statistics. Means and standard deviations for the reasoning variables and age are shown in Table 1. The usual correction for guessing (score = R - W/{k-1}, where R = number right, W = number wrong, and k = number of answer choices) was applied to all of the marker tests. On Figure Classification, where 64 items have two choices and 48 have three choices, an approximate correction was made by subtracting three-fourths of the number of errors from the number correct. A guessing correction is not needed on the JOC reasoning tests because on each the number of possible answers is very large. It should again be pointed out that the instructions for the marker tests stated that the total score would be the number correct, in order to be consistent with the instructions for other tests in the battery. The discrepancy between the instructions and the actual scoring method is of no consequence to the examinees, because the scores have been used solely for research; however, there may be some effect on the level of scores obtained on the three ETS marker tests, for which the standard instructions discourage guessing.

Each of the tests for which age norms were not available was regressed on age and age-squared. Only one test showed a relationship with age: the Inference test has a multiple correlation of .261 with age and age-squared.

Table 1
*Descriptive Statistics for Reasoning Tests (N = 615)*

| Variable | Mean | SD |
|---|---|---|
| Age | 21.64 | 5.94 |
| Age-squared | 503 | 304 |
| Inference[a] | 6.05 | 2.42 |
| Logical Reasoning | 11.51 | 4.64 |
| Letter Sets | 8.23 | 2.74 |
| Locations | 5.54 | 3.06 |
| Figure Classification | 58.6 | 19.9 |
| Inductive Reas. (Ind)[b] | 49.9 | 27.6 |
| Analytical Reas.[b] | 47.7 | 28.3 |
| Inductive Reas. (Group) | 21.64 | 3.41 |

[a] Before adjustment for age variation.
[b] Centiles by age.

The quadratic multiple-regression equation, used in making the age adjustment, is: $score' = -.742 + (.492 \times Age) - (.00773 \times Age^2)$.

A check was also made for possible sex differences on the reasoning tests. Because the females in the sample are older than the males, this check was made using the partial correlation of sex with test score, controlling for age and age-squared. There was no evidence of any difference between male and female scores: all partial correlations were statistically nonsignificant, ranging from -.07 (females superior) for Letter Sets to .06 (males superior) for Logical Reasoning.

The intercorrelation matrix for all eight reasoning tests, after removal of age-related variance on the Inference test, is shown in Table 2.

Table 2
*Intercorrelations of Reasoning Tests (N = 615)*

| | Inf | Log Reas | Let Sets | Loc | Fig Class | IR (ind) | Ana Reas | IR (gp) |
|---|---|---|---|---|---|---|---|---|
| Inference | 1.000 | .412 | .423 | .355 | .297 | .161 | .407 | .246 |
| Logical Reas. | .412 | 1.000 | .449 | .352 | .364 | .096 | .381 | .209 |
| Letter Sets | .423 | .449 | 1.000 | .386 | .350 | .183 | .361 | .245 |
| Locations | .355 | .352 | .386 | 1.000 | .373 | .200 | .336 | .220 |
| Fig. Class. | .297 | .364 | .350 | .373 | 1.000 | .189 | .362 | .198 |
| Ind. Reas. (ind) | .161 | .096 | .183 | .200 | .189 | 1.000 | .367 | .345 |
| Ana. Reas. | .407 | .381 | .361 | .336 | .362 | .367 | 1.000 | .296 |
| Ind. Reas. (gp) | .246 | .209 | .245 | .220 | .198 | .345 | .296 | 1.000 |

II. Reliability of Group Inductive Reasoning. Because an estimate of reliability will be required for the multiple-groups factor analysis, the internal-consistency reliability of Worksample 700 was calculated for the present sample. Scores on the three separately-timed parts were used to compute coefficient alpha, which is .535. This reliability is quite low for a thirty-item test, and indicates that Worksample 700 requires fundamental revisions if it is to become a reasonably accurate measure.

III. Examination of Hypothesized Factors. The multiple-groups common factor analysis, using reliabilities as communality estimates, produced the set of factor loadings (correlations between tests and factors) shown in Table 3.

Table 3
*Factor Structure Matrix:*
*Multiple-Groups Factor Analysis*

| Variable | F1 | F2 |
|---|---|---|
| Inference | .673 | .525 |
| Logical Reasoning | .815 | .568 |
| Letter Sets | .586 | .721 |
| Locations | .475 | .663 |
| Figure Classification | .444 | .665 |
| Inductive Reas. (ind.) | .173 | .279 |
| Analytical Reas. | .530 | .517 |
| Inductive Reas. (group) | .306 | .323 |
| Correlation between factors = .735 | | |

The correlation between the two factors is higher than expected. It probably is of no practical value to distinguish between two factors that correlate this highly. Table 4 shows the means of the correlations between marker tests defining a factor and between marker tests defining different factors. The within-factor correlations are only slightly higher than the between-factor correlations, giving support to the idea that the specification of two separate factors is not justified.

Table 4
*Mean Correlations Between Marker Tests*
*Defining the Same Factor or Different Factors*

| | F1 | F2 |
|---|---|---|
| F1 (Logical Reas.) | .412 | .374 |
| F2 (Induction) | .374 | .388 |

Note: Fisher's z transformation was used in calculating mean correlations.

Because it appears that a single factor can account for the intercorrelations among the marker tests, the multiple-groups factor analysis was repeated with the five marker tests defining a general reasoning factor. The correlations of the tests with this general factor are shown in Table 5. The higher loadings for Logical Reasoning and Letter Sets than for the other marker tests are probably not meaningful, because they reflect the higher estimated reliabilities for those two tests (based on internal-consistency rather than alternate-form estimates).

Table 5
*Loadings on a General Factor Defined by the Five Marker Tests*

| | F1 |
|---|---|
| Inference | .629 |
| Logical Reasoning | .720 |
| Letter Sets | .712 |
| Locations | .626 |
| Figure Classification | .613 |
| Inductive Reas. (ind.) | .251 |
| Analytical Reasoning | .560 |
| Inductive Reas. (group) | .339 |

A comparison of the residual matrices, after extraction of either one or two multiple-group factors, will indicate the degree to which the two-factor solution can better explain the intercorrelations. These residual matrices are shown in Table 6. The mean absolute value of the residuals after extracting one factor is .093, and after two factors are extracted the mean is .083, only slightly smaller. Thus it appears quite reasonable to utilize a one-factor solution.

Table 6
*Residuals After Extraction of One (Above Diagonal) or Two (Below Diagonal) Multiple-Group Factors*

| | Inf | Log Reas | Let Sets | Loc | Fig Class | IR (ind) | Ana Reas | IR (gp) |
|---|---|---|---|---|---|---|---|---|
| Inference | .194 | -.041 | -.025 | -.039 | -.089 | .003 | .055 | .033 |
| | *.135* | | | | | | | |
| Logical Reas. | *-.135* | .282 | -.064 | -.099 | -.078 | -.085 | -.022 | -.035 |
| | | *.135* | | | | | | |
| Letter Sets | *.010* | *-.010* | .234 | -.059 | -.086 | .004 | -.038 | .004 |
| | | | *.214* | | | | | |
| Locations | *.015* | *-.014* | *-.090* | .208 | -.011 | .043 | -.014 | .008 |
| | | | | *.160* | | | | |
| Figure Class. | *-.024* | *.024* | *-.123* | *-.069* | .264 | .035 | .019 | -.010 |
| | | | | | *.193* | | | |
| Ind. Reas. (ind) | *.035* | *-.035* | *-.014* | *.014* | *.000* | .817 | .226 | .260 |
| | | | | | | *.800* | | |
| Analytical Reas. | *.042* | *-.042* | *-.030* | *-.003* | *.033* | *.234* | .337 | .106 |
| | | | | | | | *.334* | |
| Ind. Reas. (gp) | *.034* | *-.034* | *.004* | *.007* | *-.011* | *.259* | *.107* | .425 |
| | | | | | | | | *.425* |

The largest residual covariances, after extraction of either one or two multiple-group factors, are those between the JOC Inductive Reasoning test (Wks. 164) and the other two JOC reasoning tests. As shown in Tables 3 and 5, Worksample 164 correlates only slightly with the reasoning abilities measured by the marker tests, despite its high reliability. This test appears, then, to be measuring a type of reasoning ability which is distinct from both the logical reasoning and induction factors found by other researchers. Furthermore, the group Inductive Reasoning test (Wks. 700) and the Analytical Reasoning test have a moderate degree of overlap with Wks. 164, as shown in Table 2. Therefore, a diagonal factor defined by Wks. 164 has been extracted from the residual matrix following extraction of the single general reasoning factor. Table 7 gives the correlations between the tests and the diagonal factor, which is uncorrelated with the general factor.

Table 7
*Loadings on Diagonal Factor Defined by Inductive Reasoning, Wks. 164*

| Variable | Loading |
|---|---|
| Inference | .004 |
| Logical Reasoning | -.094 |
| Letter Sets | .005 |
| Locations | .047 |
| Figure Classification | .038 |
| Inductive Reas. (ind) | .904 |
| Analytical Reasoning | .250 |
| Inductive Reas. (group) | .287 |

These two orthogonal factors--the general factor defined by the marker tests and the factor defined by Worksample 164--do a good job of explaining the intercorrelations among the eight tests. The mean absolute value of the elements in the residual matrix is only .047.

These results can be checked using Rao's canonical factor analysis. On the basis of the results obtained so far, it is expected that this analysis will show a strong general reasoning factor with high loadings for the marker tests and Analytical Reasoning; there may or may not be a second factor, corresponding to Inductive Reasoning Wks. 164. Results of the analysis are shown in Table 8. A statistically significant amount of covariance remains after extraction of the first factor ($\chi^2(20) = 199.4$, $p < .001$), but not after extraction of the second ($\chi^2(13) = 17.4$, $p > .10$). The left side of Table 8 shows the unrotated factor matrix, and the right side gives the loadings after a direct oblique rotation with $\delta = 0$. A plot of the tests against varimax-rotated orthogonal axes is shown in Figure 1.

Worksample 164 is clearly distinct from the cluster of marker tests. Analytical Reasoning is closer to the marker tests than to Worksample 164. The group-administered Inductive Reasoning test has only modest loadings on both factors. It is notable that there is no tendency for the marker tests to segregate into clusters interpretable as logical reasoning and induction.

Table 8
*Results of Canonical Factor Analysis*

| Variable | Unrotated F1 | Unrotated F2 | Oblique Rotation[a] F1 | Oblique Rotation[a] F2 | Iterated $h^2$ |
|---|---|---|---|---|---|
| Inference | .605 | -.160 | .626 | .289 | .39 |
| Logical Reasoning | .614 | -.293 | .669 | .197 | .46 |
| Letter Sets | .632 | -.182 | .657 | .291 | .43 |
| Locations | .566 | -.094 | .571 | .311 | .33 |
| Figure Classification | .544 | -.089 | .549 | .301 | .30 |
| Inductive Reas. (ind) | .446 | .590 | .276 | .736 | .55 |
| Analytical Reasoning | .654 | .115 | .601 | .525 | .44 |
| Inductive Reas. (group) | .441 | .236 | .363 | .471 | .25 |
| Eigenvalue | 3.183 | 1.115 | | | |

[a] Factor structure (correlations between tests and factors). The correlation between the factors is .456.

IV. Overlap of JOC Reasoning Tests with Reasoning Factors. In order to determine how well the Johnson O'Connor reasoning tests as a group measure the reasoning factor(s) defined by the marker tests, each of the reasoning factors from the multiple-group factor analysis--the general factor and the specific induction and logical reasoning factors--has separately been regressed on the three JOC tests. The results are shown in Table 9. In no case does Wks. 164 make a statistically significant contribution to the prediction of factor scores, and the contribution of Wks. 700, while statistically significant, is minor. Essentially, the Analytical Reasoning test alone indicates an examinee's standing on the general and specific reasoning factors. It does a moderately good job of measuring the general reasoning factor.

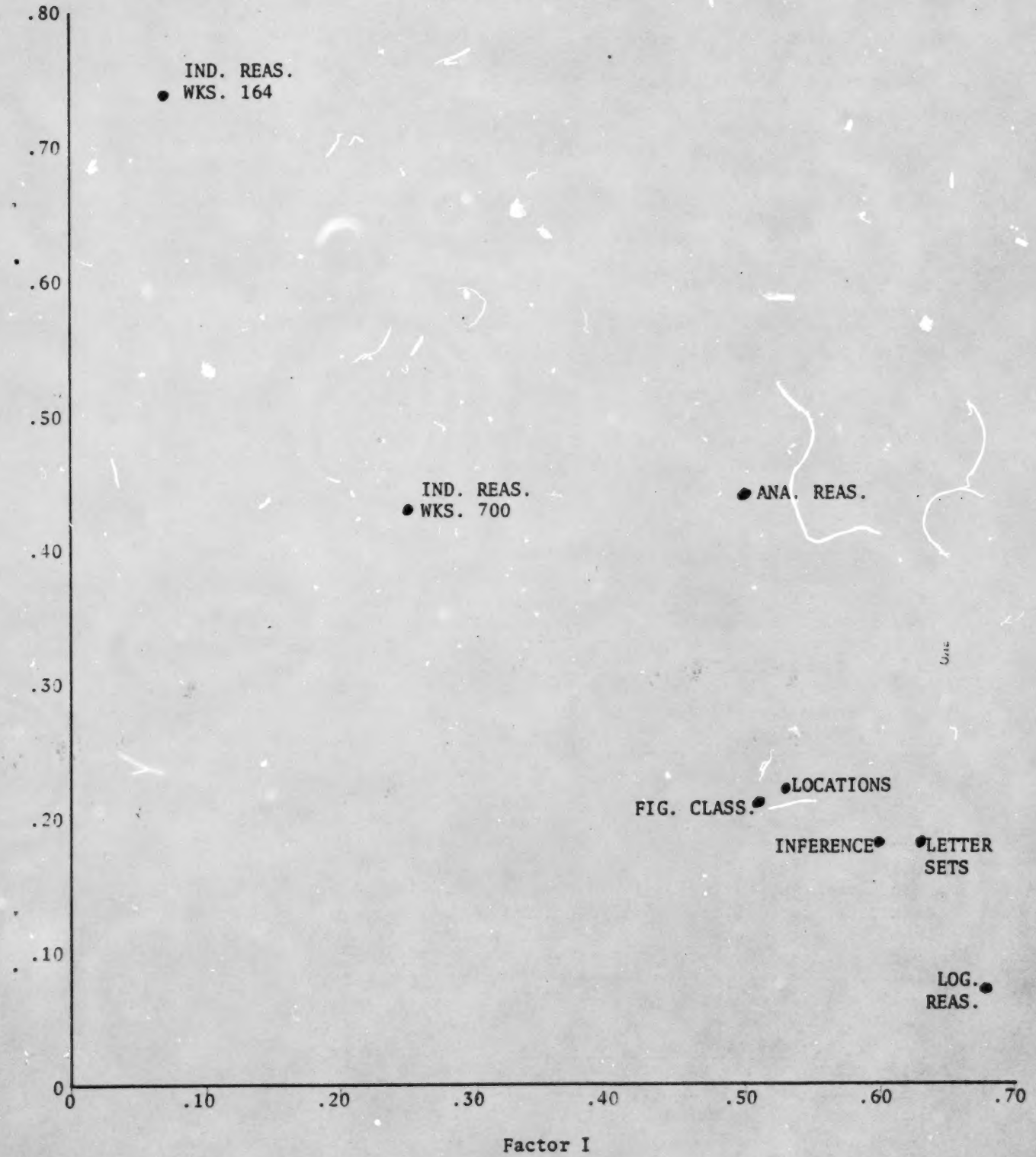
Table 9
*Multiple Regressions of Multiple-Group Factors on JOC Reasoning Tests*

| | Dependent Variable: General Reas. $R$ | General Reas. β | Induction $R$ | Induction β | Logical Reas. $R$ | Logical Reas. β |
|---|---|---|---|---|---|---|
| Analytical Reasoning | .560 | .503 | .517 | .445 | .530 | .503 |
| Inductive Reas. (group) | .588 | .189 | .547 | .172 | .552 | .182 |
| Inductive Reas. (ind) | .588 | .001 | .549 | .056 | .556 | .074 |
| Shrunken estimate of $R_p$ | .586 | | .547 | | .554 | |

V. Difference Between the Two Inductive Reasoning Tests. We have already seen that the group Inductive Reasoning test has low internal reliability and is more strongly related than Wks. 164 to the general reasoning factor. The intercorrelation of the two tests, corrected for unreliability, is only .49. In order to investigate the differences between these tests

Figure 1

*Results of Rao's Canonical Factor Analysis: Loadings on Varimax-Rotated Factors*



more fully, they have been compared in a multivariate analysis with the other aptitude tests in the JOC battery, as well as with the marker tests of reasoning. The difference between standardized scores on Worksamples 164 and 700 has been regressed on the other aptitude scores. Because there are sex differences on some of the tests, the analyses have been done separately for males and females; also, age-related score variation has been statistically removed from the tests lacking age norms, if age accounts for over one percent of test score variance. Composite scores (sums of standard scores) have been formed for structural visualization (Wiggly Block and Paper Folding) and general reasoning (the five marker tests). Results are given in Table 10.

Table 10

*Multiple Regressions of Inductive Reasoning Difference Scores (700-164) on Other Aptitude Measures*

| Females ($N$ = 246) | | | Males ($N$ = 356) | | |
|---|---|---|---|---|---|
| Variable | $R$ | $\beta$ | Variable | $R$ | $\beta$ |
| Analytical Reas. | .159 | -.156 | Writing Speed | .140 | -.124 |
| General reasoning | .185 | .210 | General reasoning | .196 | .187 |
| Structural vis. | .215 | -.165 | Number Checking | .226 | -.121 |
| Ideaphoria | .250 | -.135 | Object Observation | .249 | -.107 |

Note: only predictors which are statistically significant at the .05 level are shown. Weights are for the four-test prediction equation.

Positive beta ($\beta$) weights in the table indicate that the aptitude contributes to higher performance on Wks. 700 than on Wks. 164, and negative weights indicate the opposite. Two consistent results emerge from these regression analyses. First, as already noted, Worksample 700 has greater overlap than Wks. 164 with the general reasoning factor defined by the marker tests. Second, Wks. 164 is more strongly related to highly speeded tests such as Number Checking, Writing Speed and Ideaphoria. The stronger relationship of Analytical Reasoning to Wks. 164 than to Wks. 700 in the female sample (when general reasoning is held statistically constant) supports the interpretation that the similar scoring method on Analytical Reasoning and Wks. 164 (i.e. each item scored on completion time) contributes to their intercorrelation. The results for structural visualization and Object Observation are not replicated across sex; the zero-order correlations of these aptitudes with the difference score have opposite signs in the two samples.

In conclusion, the only differences between the group and individually-administered forms of Inductive Reasoning are that the former is more related to the general reasoning factor, and the latter is more affected by speed. It appears that the test as scored on individual item-completion times is more sensitive to speed than is the group-administered test, even though the latter has short time limits for its subsections. It is not clear why Wks. 700 should be more influenced by the general reasoning ability; perhaps this is partly attributable to the fact that Wks. 700 was administered on the same tape as four of the five marker tests.

VI. Relationship of Reasoning Tests to Other Aptitudes. The results described up to this point could have been influenced by non-reasoning abilities, such as structural visualization, graphoria, memory or vocabulary. For example, the correlation between pairs of tests could be affected by their common correlation with a third test. For this reason it is useful to compare the correlations of all of the reasoning tests with other aptitudes. This information is presented in Table 11, separately for males and females.

The strongest correlations of the marker tests are with structural visualization and English vocabulary. Figure Classification has a relatively low correlation with vocabulary, and Inference has a relatively low correlation with structural visualization. In general, however, the intercorrelations of the marker tests can be "explained" to a considerable extent by their common relationships with these two abilities. (Of course, it is an open question whether these abilities affect marker-test scores, reasoning ability affects scores on spatial and vocabulary tests, or both. Hence there is no intention here to "explain away" the correlations among the marker tests.) It is notable that Analytical Reasoning shows the same pattern of correlations with outside tests as is shown by the marker tests, whereas both forms of Inductive Reasoning are relatively independent of vocabulary and (at least among females) of spatial visualization.

To some extent, these correlations can be interpreted as demonstrating the influence of spatial visualization and English vocabulary on reasoning test performance. For example, Logical Reasoning, Inference, and Analytical Reasoning all use verbal material, and Letter Sets uses letters, so a correlation of these tests with vocabulary is to be expected. Figure Classification is the most obviously "spatial" of the reasoning tests, but Locations and Analytical Reasoning both call for the perception of two-dimensional patterns. Nevertheless, the correlations which can be explained through this type of direct examination of the tests are not a great deal higher than the correlations which have no such obvious explanation: for example, the correlations of Letter Sets and Logical Reasoning with spatial visualization, and Figure Classification and Locations with vocabulary, are substantial. This implies that the general reasoning ability common to the marker tests (and Analytical Reasoning) also influences performance on the spatial (particularly Paper Folding) and vocabulary tests. It certainly is plausible that general reasoning could aid in the solution of three-dimensional puzzles. It is less clear how reasoning would assist in answering vocabulary-test items (although it might be helpful on such a test in analyzing patterns in the sets of answer choices). In the case of both structural visualization and English vocabulary, it is possible that the underlying aptitudes are themselves correlated with reasoning ability, for any of a variety of reasons. For example, reasoning ability probably helps a person figure out the meaning of a word from context, and thereby will contribute to an enlarged vocabulary.

Whether general reasoning ability helps one superficially in the solution of vocabulary and spatial test items, or is correlated with the underlying aptitudes, it seems clear from the data in Table 11 that it would be fruitless to try to eliminate the correlation of reasoning tests with spatial and vocabulary tests by revising the reasoning tests themselves. On the other hand, some of the general reasoning tests are more independent of the influence of spatial and verbal ability than others, and might therefore be preferred as part of a test battery.

Table 11

*Correlations of Reasoning Tests with Other Aptitude Tests (Males/Females)*

| Aptitude | Logical Reas | Infer-ence | Letter Sets | Figure Class | Loca-tions | General Reas | IR (ind) | IR (group) | Ana Reas |
|---|---|---|---|---|---|---|---|---|---|
| Graphoria | .14/.08 | .19/.14 | .29/.20 | .14/.22 | .10/.22 | .24/.24 | .22/.09 | .06/.20 | .20/.11 |
| Ideaphoria | .09/.14 | .20/.26 | .22/.22 | .10/.14 | .18/.14 | .23/.25 | .18/.11 | .06/.26 | .17/.26 |
| Structural Visualization | .37/.33 | .26/.33 | .41/.47 | .47/.47 | .44/.44 | .55/.57 | .28/.13 | .35/.26 | .50/.43 |
| Wiggly Block | .28/.18 | .18/.26 | .27/.35 | .39/.33 | .34/.31 | .42/.40 | .27/.13 | .31/.31 | .42/.38 |
| Paper Folding | .38/.40 | .29/.32 | .46/.47 | .46/.48 | .44/.46 | .58/.60 | .24/.14 | .33/.15 | .49/.38 |
| Memory for Design | .27/.30 | .20/.25 | .42/.36 | .29/.34 | .39/.22 | .44/.41 | .28/.18 | .30/.18 | .38/.38 |
| Silograms | .26/.31 | .29/.33 | .30/.21 | .10/.26 | .20/.33 | .33/.40 | .18/.18 | .19/.18 | .21/.38 |
| Number Memory | .25/.25 | .16/.25 | .29/.21 | .16/.20 | .28/.21 | .32/.32 | .17/.10 | .15/.06 | .18/.26 |
| Object Observation | .13/.20 | .15/.17 | .20/.15 | .20/.20 | .20/.13 | .25/.24 | .26/.21 | .17/.18 | .31/.30 |
| Writing Speed | .11/.15 | .15/.13 | .23/.17 | .12/.18 | .13/.04 | .21/.19 | .15/.00 | -.02/.10 | .19/.14 |
| English Vocabulary | .44/.43 | .47/.48 | .40/.36 | .29/.29 | .30/.35 | .54/.54 | .10/.15 | .20/.18 | .40/.47 |

[a] Sum of standard scores on five marker tests.

Note: First correlation of each pair is for males ($N = 356$), second is for females ($N = 246$).

The Inductive Reasoning tests have little obvious connection with either spatial visualization or vocabulary. The low correlations of the Inductive Reasoning tests with vocabulary and their relatively modest correlations with structural visualization also imply that the Inductive Reasoning aptitude is of little help in doing the spatial and vocabulary tests, and is relatively uncorrelated with the underlying aptitudes.

The other correlations in Table 11 do not require a great deal of comment. The Memory for Design test correlates about .5 with spatial visualization, and the Silograms test correlates about .4 with English vocabulary, so the correlations of these tests with the reasoning tests can be explained through their common relationship with spatial and verbal ability. The other aptitudes do not influence the reasoning tests to any substantial extent.

## CONCLUSIONS

I. The hypothesized factors of induction and logical reasoning are found to be highly overlapping, correlating .73 with each other. This result is not inconsistent with past findings, because reasoning tests have often been found to load on more than one reasoning factor. However, the outcome does suggest that a distinction between these two factors is of limited usefulness. A single general factor of reasoning does a good job of explaining the correlations among the marker tests for induction and logical reasoning.

II. Inductive Reasoning (Wks. 164) is largely independent of the general reasoning factor. It seems primarily to measure quickness in seeing relationships among items, whereas the marker tests call for a substantial amount of hypothesis testing in addition to, or instead of, the perception of relationships. Worksample 164 is also considerably more independent of spatial visualization and English vocabulary abilities than are the other reasoning tests.

III. Analytical Reasoning is a good measure of the general reasoning factor. It correlates more highly with Worksample 164 than the marker tests do, but this may be because Analytical Reasoning and Wks. 164 are both scored on item completion times, and may therefore be more sensitive to speed than are the tests scored on number of items answered correctly in a time limit.

IV. The group Inductive Reasoning test (Wks. 700) has inadequate reliability (.54), and correlates moderately both with Wks. 164 and with the general reasoning factor. It is likely that this type of item, designed for use in a speed test where errors are infrequent, is not well suited to a time-limit administration. The test might prove satisfactory if administered in the same manner as Wks. 164. To obtain high reliability with a group administration, it might be necessary either to greatly increase the number of items, or make the items more difficult. The latter course would almost certainly reduce the correlation with Wks. 164, because it would turn the test into a power measure rather than a speed measure.

V. Assuming that Inductive Reasoning (Wks. 164) will be part of the Johnson O'Connor Research Foundation test battery, one of the marker tests might be better than Analytical Reasoning as a second reasoning test. The marker tests are more independent of Worksample 164, and two of them (Figure Classification and Locations) have lower correlations with English vocabulary,

without correlating any more highly with structural visualization. On the other hand, if it is the common scoring technique that is responsible for the greater overlap of Analytical Reasoning with Worksample 164, then the group version of Analytical Reasoning, which is administered as a time-limit test, may be just as independent of Worksample 164 as the marker tests are.

In selecting a test, validity is a more important consideration than statistical independence from other tests, especially when the structure of abilities is as poorly understood as it is in the reasoning domain. The Analytical Reasoning test has recently been found to have good predictive validity for grades in high school mathematics and scores on the Math section of the Scholastic Aptitude Test, even after partialling out English vocabulary (Daniel, 1982). It is possible that scoring a test on item-completion times yields greater validity than scoring it on a time-limit basis; speed may be an important indicator of reasoning ability for a variety of item types, and the timing of individual items may yield a score that is more sensitive to speed. Therefore, replacing Analytical Reasoning with a test that has a lower correlation with Worksample 164 will not necessarily represent an improvement of the battery.

VI. The correlations between the tests of general reasoning (including Analytical Reasoning) and the tests of spatial ability and English vocabulary appear to be caused, to a considerable extent, by factors other than verbal or spatial elements in the reasoning tests. Reasoning ability may help in the taking of the verbal and spatial tests, or the reasoning aptitude may be correlated with the underlying verbal and spatial abilities. Therefore, it is unlikely that the reasoning tests could be modified in a way that would make them independent of spatial visualization or English vocabulary knowledge, without detracting from their validity as measures of reasoning. Of course, some of the general reasoning tests are influenced more than others by spatial or verbal ability because of their specific content, and preference might reasonably be given to those tests that are more independent of these factors. It is possible that the spatial tests could be revised to reduce their correlations with general reasoning; it is apparent from the present study that some spatial tests (e.g., Wiggly Block) are less influenced by reasoning ability than are others (e.g., Paper Folding), even after adjusting for differences in reliability.

## REFERENCES

Adkins, D.C. & Lyerly, S.B. Factor analysis of reasoning tests. Chapel Hill, N.C.: University of North Carolina Press. 1952.

Bowker, R. & Moore, G. Demographic information concerning Foundation examinees (Stat. Bull. 1981-37). Boston: Johnson O'Connor Research Foundation. 1981.

Daniel, M. Inductive Reasoning Worksample 164 test manual. Boston: Johnson O'Connor Research Foundation. 1978.

Daniel, M. Form EI of the Inductive Reasoning test, Worksample 164 (Tech. Rep. 867). Boston: Johnson O'Connor Research Foundation. 1979.

Daniel, M. Relationship of aptitudes to performance in high school mathematics (Tech. Rep. 1982-2). Boston: Johnson O'Connor Research Foundation. 1982.

Daniel, M. & Kling, J. Reliability of Worksample 244 CM, Analytical Reasoning (Stat. Bull. 1974-13). Boston: Johnson O'Connor Research Foundation. 1974.

Ekstrom, R.B., French, J.W., & Harman, H.H. Manual for kit of factor-referenced cognitive tests. Princeton: Educational Testing Service. 1976.

Gorsuch, R.L. Factor analysis. Philadelphia: Saunders. 1974.

Green, R.F., Guilford, J.P., Christensen, P.R. & Comrey, A.L. A factor-analytic study of reasoning abilities. Psychometrika, 1953, 18, 135-160.

Guilford, J.P., Green, R.F., Christensen, P.R., Hertzka, A.F. & Kettner, N.W. A factor-analytic study of Navy reasoning tests with the Air Force Aircrew Classification Battery. Educational and Psychological Measurement, 1954, 14, 301-325.

Guilford, J.P. & Lacey, J.I. (Eds.). Printed Classification Tests. AAF Aviation Psychology Research Program Reports: Report 5. Washington, D.C.: G.P.O. 1947.

Hakstian, A.R. & Cattell, R.B. Manual for the Comprehensive Ability Battery (CAB). Champaign, Illinois: Institute for Personality and Ability Testing. 1976.

Hertzka, A.F. & Guilford, J.P. Logical Reasoning. Orange, California: Sheridan Psychological Services, Inc. 1955.

Horn, J.L. & Cattell, R.B. Whimsy and misunderstandings of Gf-Gc theory: A comment on Guilford. Psychological Bulletin, 1982, 91, 623-633.

Michael, W.B., Zimmerman, W.S. & Guilford, J.P. An investigation of the nature of the spatial relations and visualization factors in two high school samples. Educational and Psychological Measurement, 1951, 11, 561-577.

Northrup, L.C. The definition and measurement of reasoning (Technical Memorandum 77-10, Personnel Measurement Research and Development Center). Washington, D.C.: Government Printing Office. 1977.

Pemberton, C. The closure factors related to other cognitive processes. Psychometrika, 1952, 17, 267-288.

Thurstone, L.L. Primary mental abilities. Psychometric Monographs, 1938, No. 1.

Thurstone, L.L. & Thurstone, T.G. Factorial studies of intelligence. Psychometric Monographs, 1941, No. 2.

Trafton, R.S. Construct validity of the Wiggly Block and Paper Folding tests: A factor-referenced multivariate study (Tech. Rep. 1981-2). Boston: Johnson O'Connor Research Foundation. 1981.